
Introduction

Table of Contents

Data Preparation	1
Validation	1

The USPS Handwritten Digit database is a set of handwriting data containing 1100 examples of the digits 0-9. To process and learn this data, a linear regression algorithm is implemented with

Data Preparation

To prepare for training, data is pulled from the USPS handwriting set and processed. `getfeatures()` is called to find the symmetry and intensity of each data point.

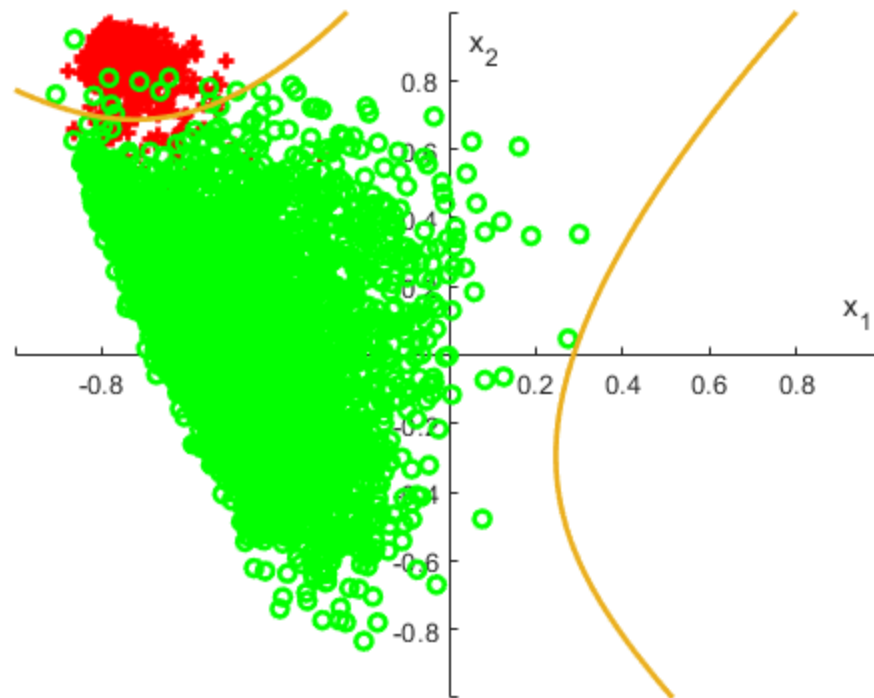
```
load('usps_modified.mat');
[xd,yd] = getfeatures(data);
x = nl_transform(xd);
for i = 501:height(yd)
    yd(i) = -1;
end
```

Validation

10-fold validation was performed on the data set, and produced positive results. Classification error and squared error were just under 3% for all values of lambda and for both the in and out samples. For all four of those cases, a lambda of 0.01 proved to be the most accurate, though only by about 0.1% in each case.

```
fold = 10;
lam = [0.001; 0.01; 0.1; 0.25; 0.5; 0.75];
ein = zeros(height(lam),1);
eout = zeros(height(lam),1);
for i = 1:height(lam)
    [tce,tse,vce,vse,~] = validation(x,yd,fold,lam(i));
    ein(i) = mean(tse);
    eout(i) = mean(vse);
end
ein_avg = mean(ein);
eout_avg = mean(eout);

% plot for illustration
[~,~,~,~,g] = validation(x,yd,fold,0.01);
h_plot(xd, yd, g);
```



Published with MATLAB® R2021a